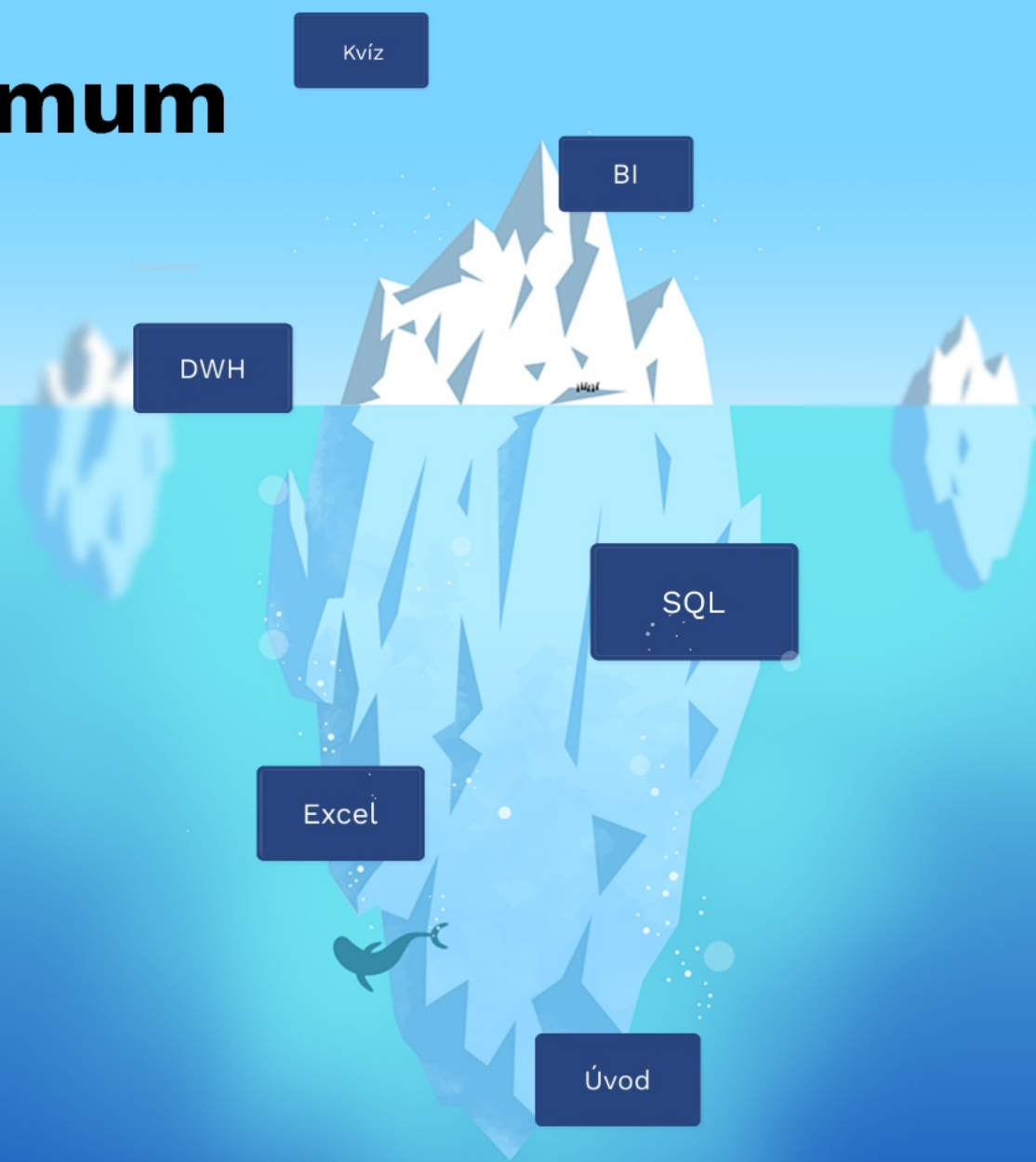


# Datové minimum



Kvíz

BI

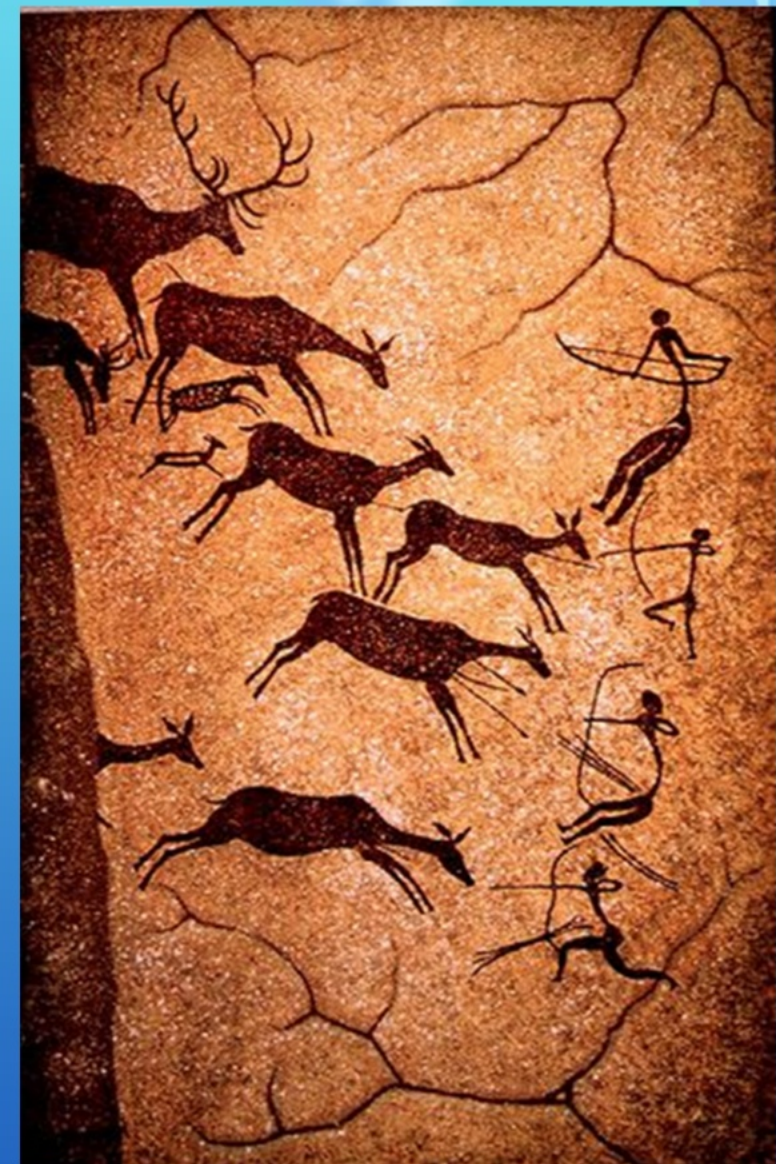
DWH

SQL

Excel

Úvod





Lovecká scéna, Los Caballos, Španělsko  
10 000-9000 př.nl

**Již naši předci publikovali data o životním prostředí.**

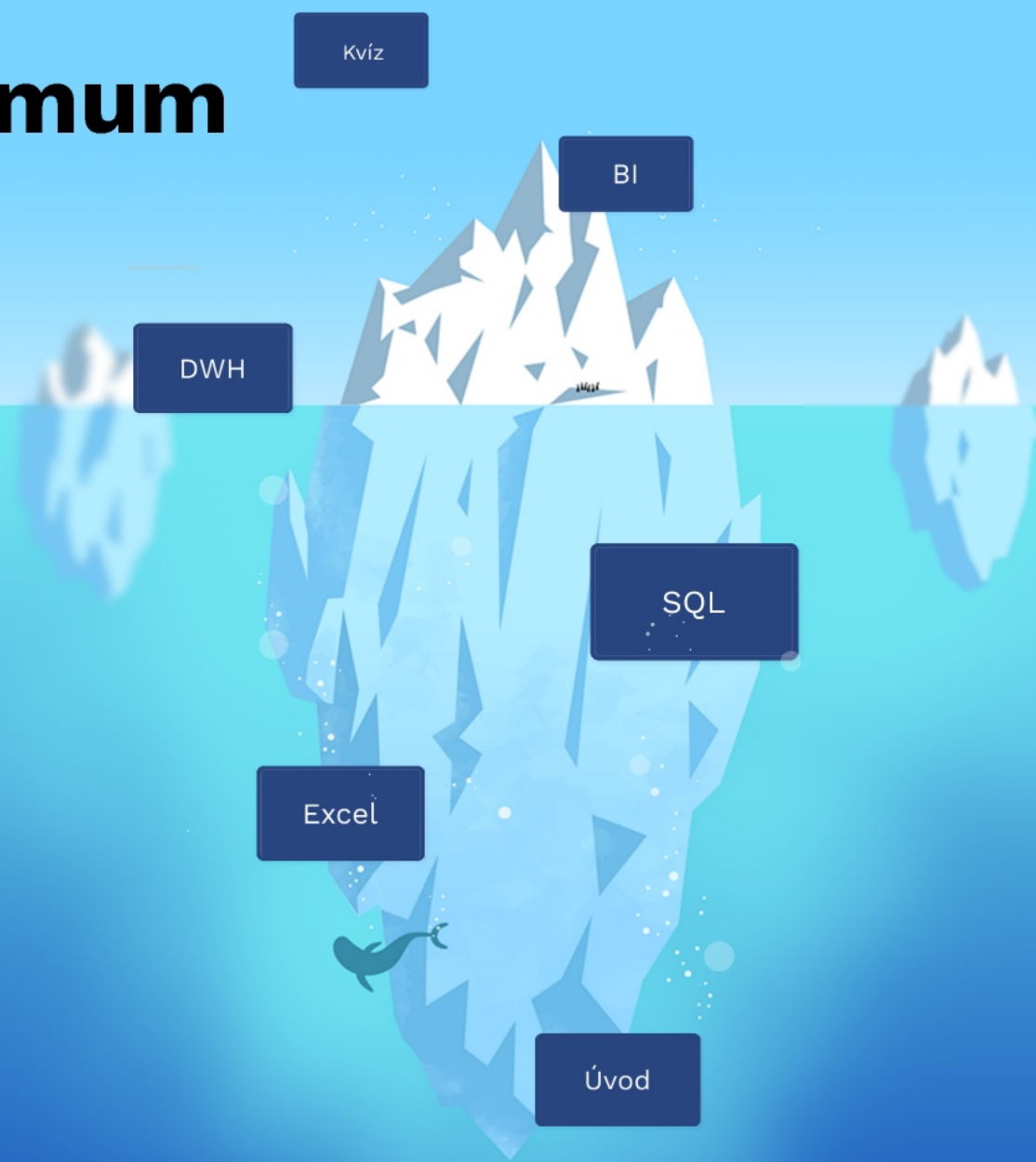
**Ctíme tradice a snažíme se v jejich práci pokračovat.**

**Jak ale zajistit, aby naše data byly dostupné všem a co nejrychleji? Vejdeme se všichni do jeskyně?**

**Bylo potřeba najít lepší způsob...**

**Digitalizaci**

# Datové minimum



Kvíz

BI

DWH

SQL

Excel

Úvod

# Excel

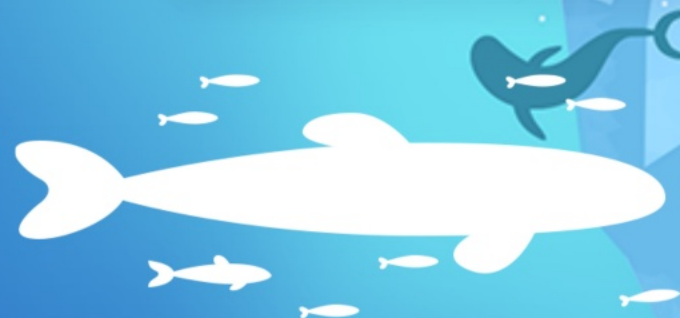
Microsoft Excel je i nadále první volbou, pokud jde o rychlou evidenci dat a jejich koncovou analýzu. Jedná se o velmi všestranný nástroj, který může dělat všechno od vytvoření jednoduchého rozpočtu až po řešení komplexních statistik.

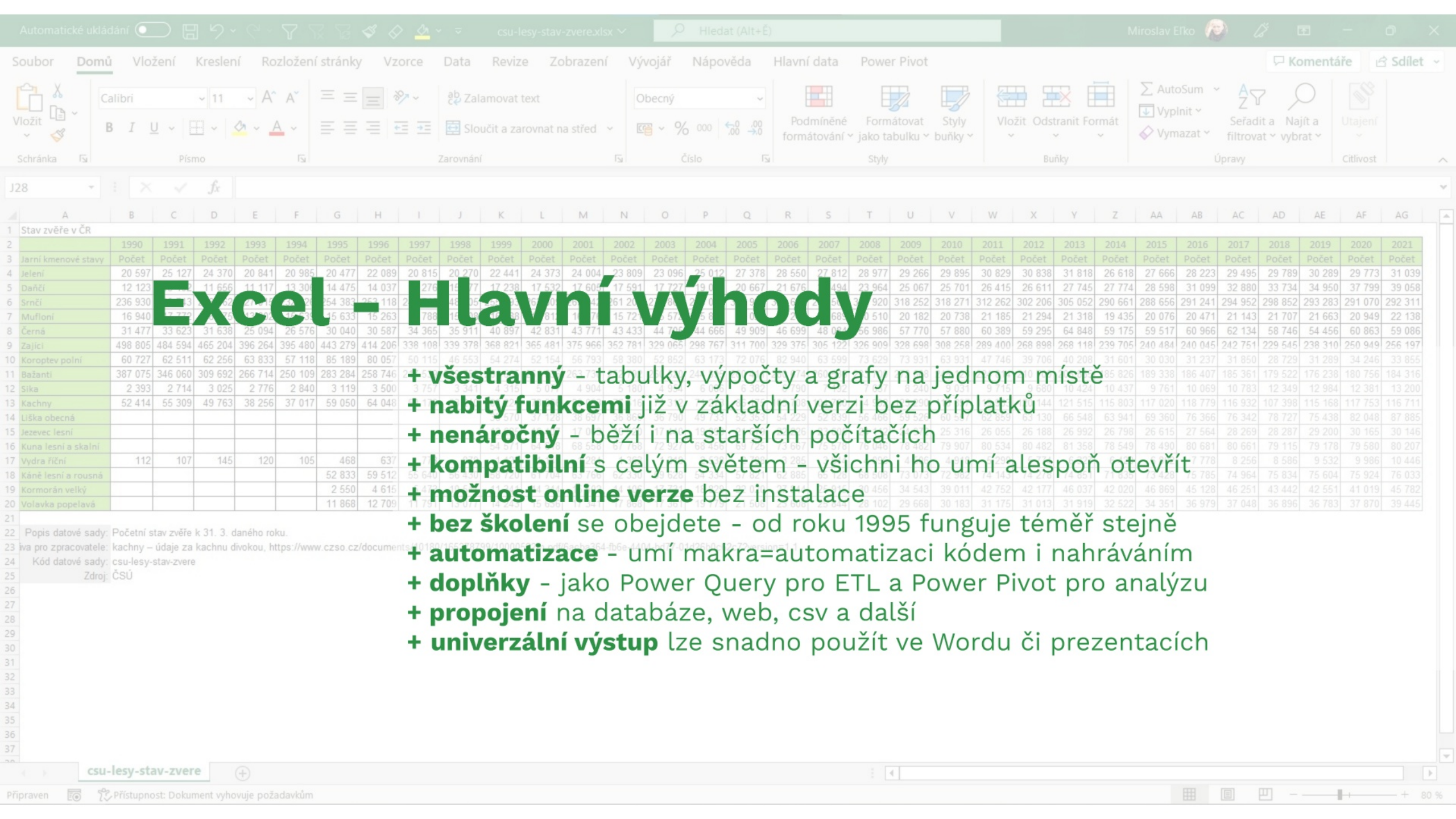
Tato vlastnost je nesporně kladná při jednoduchých a jednorázových analýzách, které neopustí náš hard-disk.

Stačí to ale?

+

-





# Excel - Hlavní výhody

- + **všestranný** - tabulky, výpočty a grafy na jednom místě
- + **nabitý funkcemi** již v základní verzi bez příplatků
- + **nenáročný** - běží i na starších počítačích
- + **kompatibilní** s celým světem - všichni ho umí alespoň otevřít
- + **možnost online verze** bez instalace
- + **bez školení** se obejdete - od roku 1995 funguje téměř stejně
- + **automatizace** - umí makra=automatizaci kódem i nahráváním
- + **doplňky** - jako Power Query pro ETL a Power Pivot pro analýzu
- + **propojení** na databáze, web, csv a další
- + **univerzální výstup** lze snadno použít ve Wordu či prezentacích

# Excel

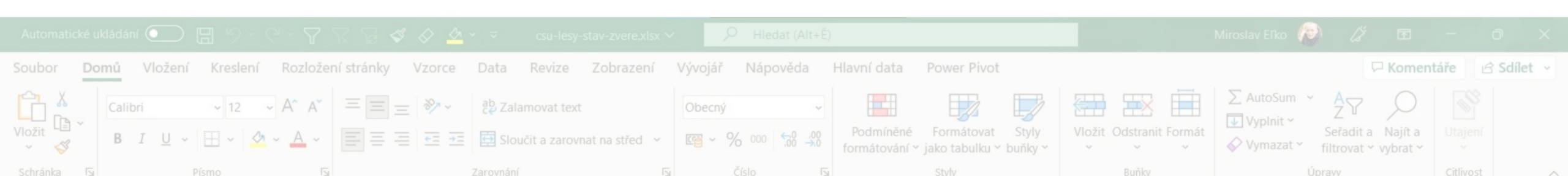
Microsoft Excel je i nadále první volbou, pokud jde o rychlou evidenci dat a jejich koncovou analýzu. Jedná se o velmi všestranný nástroj, který může dělat všechno od vytvoření jednoduchého rozpočtu až po řešení komplexních statistik.

Tato vlastnost je nesporně kladná při jednoduchých a jednorázových analýzách, které neopustí náš hard-disk.

Stačí to ale?

+

-



Name	Code	Iri položky z referenčního	Definice	Název nadřazeného	Kód nadřazené	Název nadřazené	Vztah k nadřazené	Zobrazovaný název	Platnost od	Platnost do	Alternativní název 1	Alternativní název 2	Alternativní název 3	EN název položky	EN alternativní název 1	EN alternativní název 2
Bažanti	bažanti								01.01.1901	09.09.9999						
Černá	cerna								01.01.1901	09.09.9999						
Daňčí	danci								01.01.1901	09.09.9999						
Jelení	jeleni								01.01.1901	09.09.9999						
Jezevec lesní	jezevec-lesni								01.01.1901	09.09.9999						
Kachny	kachny								01.01.1901	09.09.9999						
Káně lesní a rousná	kane-lesni-a-rousna								01.01.1901	09.09.9999						
Kormorán velký	kormoran-velky								01.01.1901	09.09.9999						
Koroptev polní	koroptev-polni								01.01.1901	09.09.9999	Koroptve					
Kuna lesní a skalní	kuna-lesni-a-skalni								01.01.1901	09.09.9999						
Liška obecná	liska-obecna								01.01.1901	09.09.9999						
Mufloní	muffloni								01.01.1901	09.09.9999						
neurčeno	neurceno								01.01.1901	09.09.9999						
Sika	sika								01.01.1901	09.09.9999						
Srnčí	srnci								01.01.1901	09.09.9999						
Volavka popelavá	volavka-popelava								01.01.1901	09.09.9999						
Vydra říční	vydra-ricni								01.01.1901	09.09.9999						
Zajíci	zajici								01.01.1901	09.09.9999						

# Excel - Hlavní nevýhody

- omezené množství dat
- výkon a stabilita
- hledání chyb není triviální
- porovnávání verzí je náročné
- aktualizace reportu je často manuální práce
- sdílení souborů je nespolehlivé
- omezená online verze nepodporuje všechny funkce a makra
- nedostatečná bezpečnost dat
- publikace výstupů vyžaduje spuštění aplikace

# Excel

Microsoft Excel je i nadále první volbou, pokud jde o rychlou evidenci dat a jejich koncovou analýzu. Jedná se o velmi všestranný nástroj, který může dělat všechno od vytvoření jednoduchého rozpočtu až po řešení komplexních statistik.

Tato vlastnost je nesporně kladná při jednoduchých a jednorázových analýzách, které neopustí náš hard-disk.

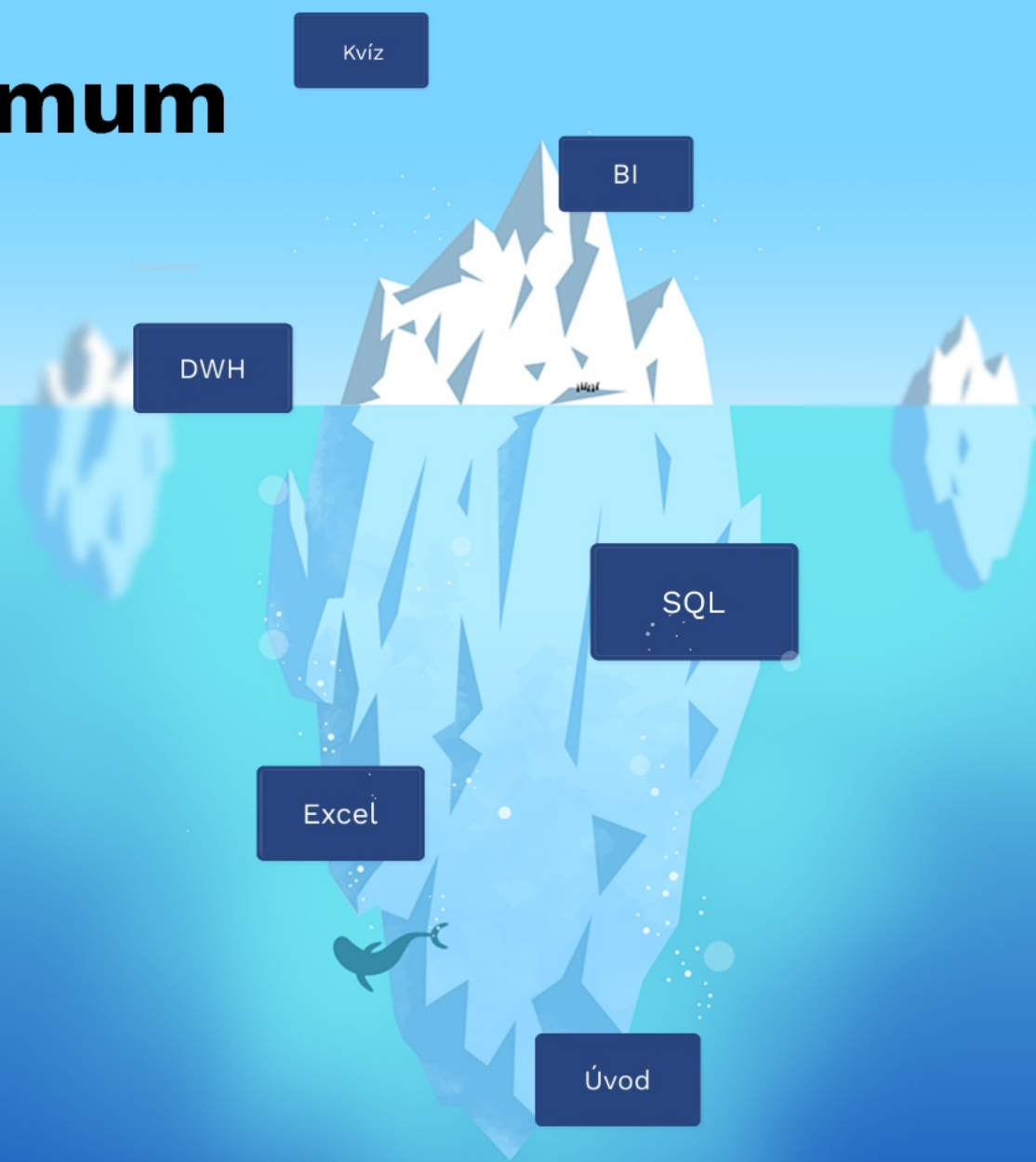
Stačí to ale?

+

-



# Datové minimum



Kvíz

BI

DWH

SQL

Excel

Úvod

# SQL

SQL (Structured Query Language/Strukturovaný dotazovací jazyk) je nástroj pro vytvoření databáze a manipulaci s daty. Jde o odpověď na velké množství dat, které již v Excelu či jiném tabulkovém editoru není efektivní spravovat.

Databáze obecně je prostor či nástroj k uchování a organizaci nějakých dat. Databází je mnoho druhů, ale my se budeme zabírat pouze nejpoužívanějším typem - databázemi relačními. Relační databáze uchovávají data v tabulkách, které jsou propojené a umožňují nám s nimi pohodlně manipulovat.

+

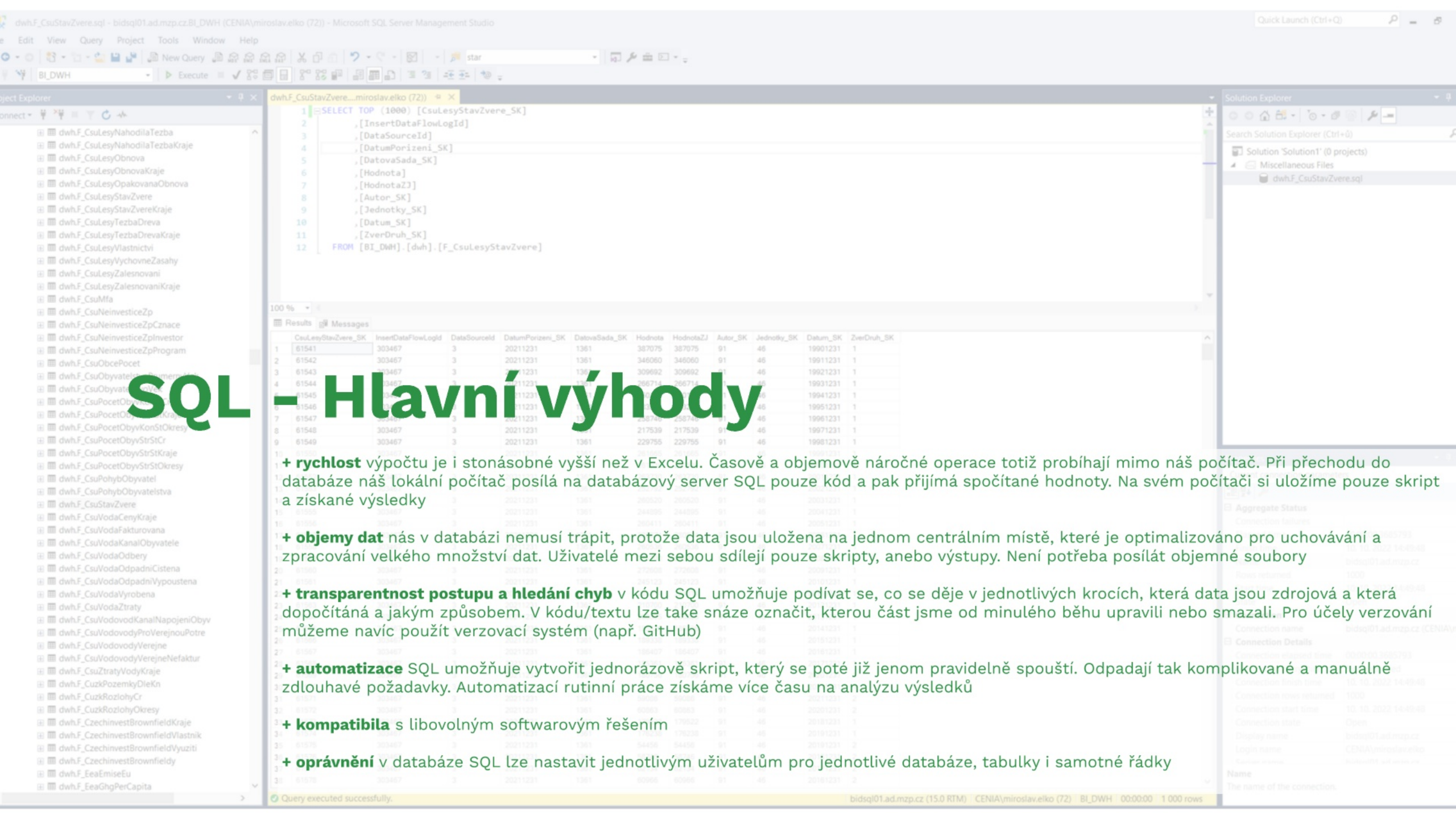
-

Software

Struktura

Pravidla

Video  
ukázka



# SQL – Hlavní výhody

- + **rychlost** výpočtu je i stonásobně vyšší než v Excelu. Časově a objemově náročné operace totiž probíhají mimo náš počítač. Při přechodu do databáze náš lokální počítač posílá na databázový server SQL pouze kód a pak přijímá spočítané hodnoty. Na svém počítači si uložíme pouze skript a získané výsledky
- + **objemy dat** nás v databázi nemusí trápit, protože data jsou uložena na jednom centrálním místě, které je optimalizováno pro uchování a zpracování velkého množství dat. Uživatelé mezi sebou sdílejí pouze skripty, anebo výstupy. Není potřeba posílat objemné soubory
- + **transparentnost postupu a hledání chyb** v kódu SQL umožňuje podívat se, co se děje v jednotlivých krocích, která data jsou zdrojová a která dopočítána a jakým způsobem. V kódu/textu lze také snáze označit, kterou část jsme od minulého běhu upravili nebo smazali. Pro účely verzování můžeme navíc použít verzovací systém (např. GitHub)
- + **automatizace** SQL umožňuje vytvořit jednorázový skript, který se poté již jenom pravidelně spouští. Odpadají tak komplikované a manuálně zdlouhavé požadavky. Automatizací rutinní práce získáme více času na analýzu výsledků
- + **kompatibilita** s libovolným softwarovým řešením
- + **oprávnění** v databáze SQL lze nastavit jednotlivým uživatelům pro jednotlivé databáze, tabulky i samotné řádky

# SQL

SQL (Structured Query Language/Strukturovaný dotazovací jazyk) je nástroj pro vytvoření databáze a manipulaci s daty. Jde o odpověď na velké množství dat, které již v Excelu či jiném tabulkovém editoru není efektivní spravovat.

Databáze obecně je prostor či nástroj k uchování a organizaci nějakých dat. Databází je mnoho druhů, ale my se budeme zabírat pouze nejpoužívanějším typem - databázemi relačními. Relační databáze uchovávají data v tabulkách, které jsou propojené a umožňují nám s nimi pohodlně manipulovat.

+

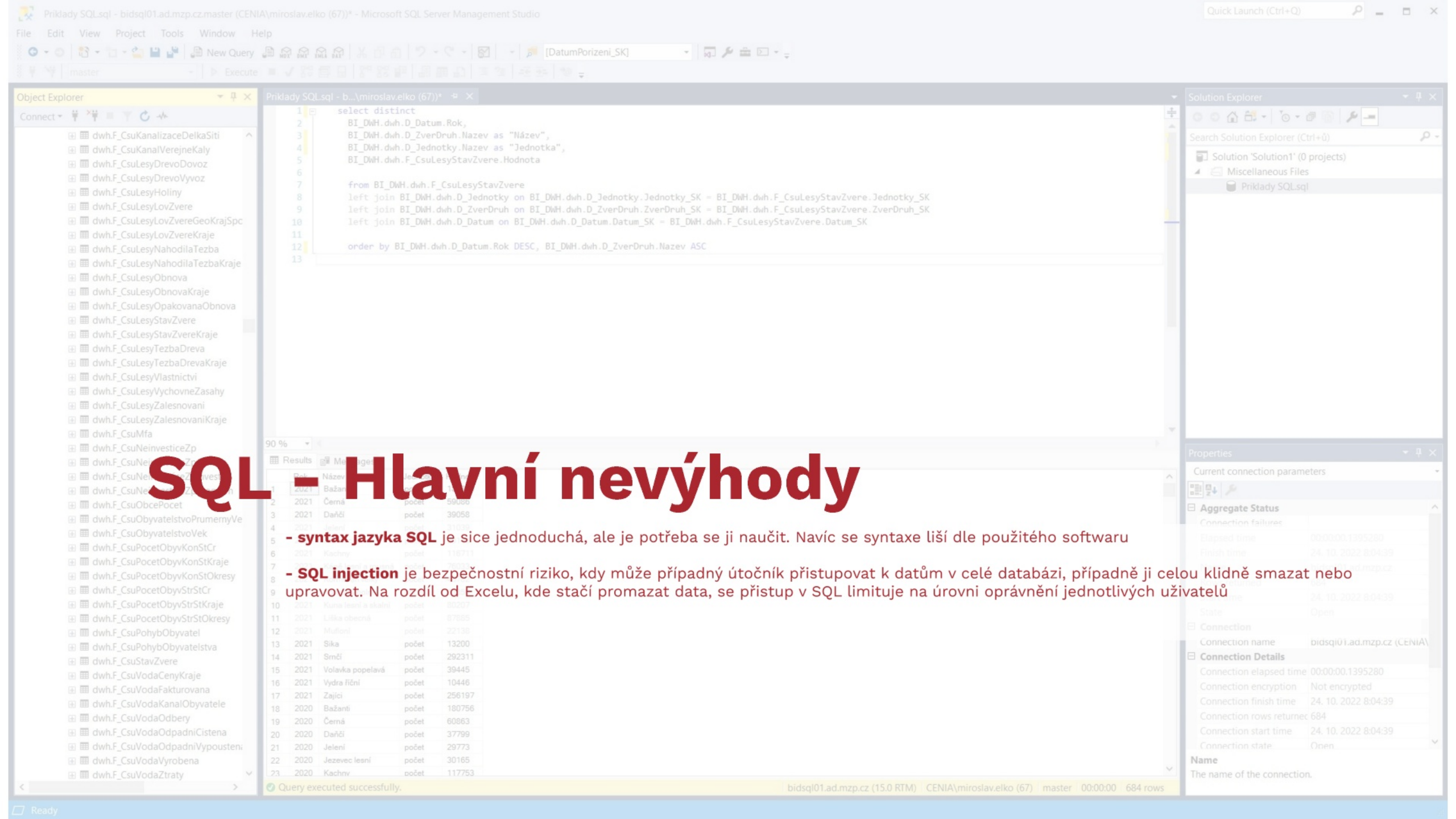
-

Software

Struktura

Pravidla

Video  
ukázka



# SQL - Hlavní nevýhody

- **syntax jazyka SQL** je sice jednoduchá, ale je potřeba se ji naučit. Navíc se syntaxe liší dle použitého softwaru
- **SQL injection** je bezpečnostní riziko, kdy může případný útočník přistupovat k datům v celé databázi, případně ji celou klidně smazat nebo upravovat. Na rozdíl od Excelu, kde stačí promazat data, se přístup v SQL limituje na úrovni oprávnění jednotlivých uživatelů

```
1 select distinct
2 BI_DWH.dwh.D_Datum.Rok,
3 BI_DWH.dwh.D_ZverDruh.Nazev as "Název",
4 BI_DWH.dwh.D_Jednotky.Nazev as "Jednotka",
5 BI_DWH.dwh.F_CsuLesyStavZvere.Hodnota
6
7 from BI_DWH.dwh.F_CsuLesyStavZvere
8 left join BI_DWH.dwh.D_Jednotky on BI_DWH.dwh.D_Jednotky.Jednotky_SK = BI_DWH.dwh.F_CsuLesyStavZvere.Jednotky_SK
9 left join BI_DWH.dwh.D_ZverDruh on BI_DWH.dwh.D_ZverDruh.ZverDruh_SK = BI_DWH.dwh.F_CsuLesyStavZvere.ZverDruh_SK
10 left join BI_DWH.dwh.D_Datum on BI_DWH.dwh.D_Datum.Datum_SK = BI_DWH.dwh.F_CsuLesyStavZvere.Datum_SK
11
12 order by BI_DWH.dwh.D_Datum.Rok DESC, BI_DWH.dwh.D_ZverDruh.Nazev ASC
13
```

Rok	Název	Jednotka	počet	Hodnota
2021	Bažant	počet	59080	
2021	Černá	počet	39058	
2021	Daňdí	počet	21736	
2021	Kachny	počet	116711	
2021	Kůna lesní a skatní	počet	80207	
2021	Lilka obecná	počet	67685	
2021	Mulloni	počet	22138	
2021	Sika	počet	13200	
2021	Směi	počet	292311	
2021	Volevka popelavá	počet	39445	
2021	Vydra říční	počet	10446	
2021	Zajici	počet	256197	
2020	Bažanti	počet	180756	
2020	Černá	počet	60863	
2020	Daňdí	počet	37799	
2020	Jeleni	počet	29773	
2020	Jezevec lesní	počet	30165	
2020	Kachny	počet	117753	

Query executed successfully.

# SQL

SQL (Structured Query Language/Strukturovaný dotazovací jazyk) je nástroj pro vytvoření databáze a manipulaci s daty. Jde o odpověď na velké množství dat, které již v Excelu či jiném tabulkovém editoru není efektivní spravovat.

Databáze obecně je prostor či nástroj k uchování a organizaci nějakých dat. Databází je mnoho druhů, ale my se budeme zabírat pouze nejpoužívanějším typem - databázemi relačními. Relační databáze uchovávají data v tabulkách, které jsou propojené a umožňují nám s nimi pohodlně manipulovat.

+

-

Software

Struktura

Pravidla

Video  
ukázka

# Software pro SQL

Existuje několik různých databázových serverů. Jsou to programy či služby, které umožňují manipulaci s daty. Nejpoužívanější jsou databázové servery MySQL (oblíbený open source), dále pak MS SQL, PostgreSQL, Oracle či Firebird. Všechny umí jazyk SQL, základ je stejný, ale v některých (mnohdy podstatných) detailech se tyto databáze a i jejich verze jazyka SQL liší.

# SQL

SQL (Structured Query Language/Strukturovaný dotazovací jazyk) je nástroj pro vytvoření databáze a manipulaci s daty. Jde o odpověď na velké množství dat, které již v Excelu či jiném tabulkovém editoru není efektivní spravovat.

Databáze obecně je prostor či nástroj k uchování a organizaci nějakých dat. Databází je mnoho druhů, ale my se budeme zabírat pouze nejpoužívanějším typem - databázemi relačními. Relační databáze uchovávají data v tabulkách, které jsou propojené a umožňují nám s nimi pohodlně manipulovat.

+

-

Software

Struktura

Pravidla

Video  
ukázka



# Struktura

## Základní struktura jazyka SQL:

**SELECT** (Povinný - Definuje obsah, který chceme zobrazit)

**FROM** (Povinný - Definuje zdroj tabulek pro tento obsah)

**WHERE** (Nepovinný - Určuje kritéria pro obsah)

**ORDER BY** (Nepovinný - Určuje, jak seřadit výsledky)

**GROUP BY** (Nepovinný - Určuje, jak seskupit výsledky)

# SQL

SQL (Structured Query Language/Strukturovaný dotazovací jazyk) je nástroj pro vytvoření databáze a manipulaci s daty. Jde o odpověď na velké množství dat, které již v Excelu či jiném tabulkovém editoru není efektivní spravovat.

Databáze obecně je prostor či nástroj k uchování a organizaci nějakých dat. Databází je mnoho druhů, ale my se budeme zabírat pouze nejpoužívanějším typem - databázemi relačními. Relační databáze uchovávají data v tabulkách, které jsou propojené a umožňují nám s nimi pohodlně manipulovat.

+

-

Software

Struktura

Pravidla

Video  
ukázka

# Pravidla

Každý sloupec musí mít svůj datový typ, který přesně určuje typ dat, které v daném sloupci budou. Nejběžnější datové typy jsou **varchar** (textová hodnota), **int** (celé číslo), **bit** (hodnota Boolean) a **datetime** (datum a čas).

Ve všech databázích existuje ještě speciální hodnota **NULL**, která znamená něco ve smyslu nevyplněno. Můžeme ji pro konkrétní sloupce v tabulce povolit, čímž řekneme, že údaj nemusí být vyplněn.

Každá tabulka by měla mít tzv. **primární klíč** - je to speciální sloupec, který slouží k jednoznačné identifikaci záznamu v tabulce. Žádné dva řádky nesmí mít v tomto sloupci stejnou hodnotu, každý řádek musí mít hodnotu primárního klíče jedinečnou (unikátní) v rámci tabulky.

# SQL

SQL (Structured Query Language/Strukturovaný dotazovací jazyk) je nástroj pro vytvoření databáze a manipulaci s daty. Jde o odpověď na velké množství dat, které již v Excelu či jiném tabulkovém editoru není efektivní spravovat.

Databáze obecně je prostor či nástroj k uchování a organizaci nějakých dat. Databází je mnoho druhů, ale my se budeme zabírat pouze nejpoužívanějším typem - databázemi relačními. Relační databáze uchovávají data v tabulkách, které jsou propojené a umožňují nám s nimi pohodlně manipulovat.

+

-

Software

Struktura

Pravidla

Video  
ukázka

# Video ukázka

Internet nabízí mnoho návodu a tipů, jak pracovat s SQL.

Skvělým průvodcem pro začátečníky v SQL je například Michael Zeidler, který má na Youtube ke dnešku již 15 epizod věnovaných této tematice.

Pojďme se podívat na jednu z nich:  
SQL Kurz Pro Začátečníky 7.



# ZÁKLADY SQL

Vyhledávání informací



# SQL

SQL (Structured Query Language/Strukturovaný dotazovací jazyk) je nástroj pro vytvoření databáze a manipulaci s daty. Jde o odpověď na velké množství dat, které již v Excelu či jiném tabulkovém editoru není efektivní spravovat.

Databáze obecně je prostor či nástroj k uchování a organizaci nějakých dat. Databází je mnoho druhů, ale my se budeme zabírat pouze nejpoužívanějším typem - databázemi relačními. Relační databáze uchovávají data v tabulkách, které jsou propojené a umožňují nám s nimi pohodlně manipulovat.

+

-

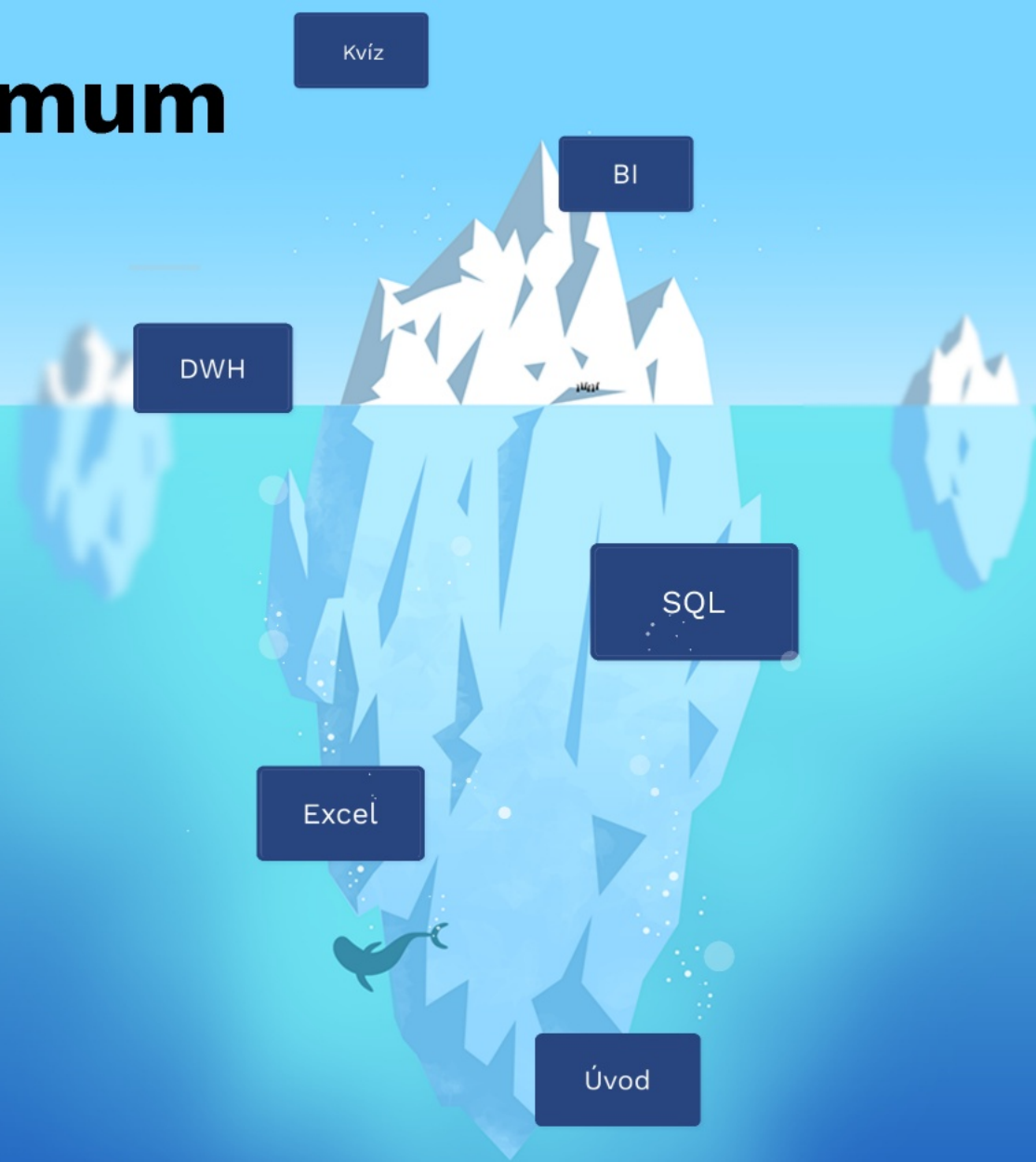
Software

Struktura

Pravidla

Video  
ukázka

# Datové minimum



Kvíz

BI

DWH

SQL

Excel

Úvod



# DWH (Data warehouse)

Databáze jsou typicky klasifikované jako relační (SQL) nebo NoSQL, transakční (OLTP), analytické (OLAP) nebo hybridní (HTAP).

Pokusy o vytvoření jednotných databází pro všechna data v rámci podniku jsou označovány jako data lakes (datová jezera), pokud se data ponechají v původním formátu. Nebo jako datové sklady, pokud se údaje převedou do společného formátu a schématu. Podmnožiny datového skladu (data warehouse) se pak nazývají datová tržiště (data mart). V rámci tohoto e-learningu nás zajímají relační databáze.

Centralizace  
dat

Čištota dat

Jak se to  
dělá?

Jak to  
funguje?

Využití

# Centralizace dat

Když se řekne sklad, každý hned ví, co si představit. Obrovskou plochu s regály až do výše několika metrů. Co když ale dodáme, že ten sklad je datový?

Objemy zpracovávaných a uchovávaných dat rostou exponenciální rychlostí. Největším problémem dneška není ani objem ani struktura dat, ale to, že se nacházejí na desítkách nejrůznějších míst. Organizace s více než deseti paralelně fungujícími informačními systémy nejsou žádnou výjimkou. Pro běžně velkou společnost je však podobné „rozházení“ dat obrovskou komplikací.

Proto je důležité data co nejvíce centralizovat a sjednotit. Není potřeba, aby zákazník Jan Novák měl v deseti různých systémech vytvořenou vlastní zákaznickou kartu s různě aktuálními adresami, telefonny a údaji. Postačí nám jedno jediné místo, kde to půjde mnohem snáz udržet ve správné podobě a aktuální. Tím místem je datový sklad, tedy databázový server, který centralizuje data z nejrůznějších systémů – ať už ta strukturovaná, či nestrukturovaná – a mezi těmito systémy také zprostředkovává jejich výměnu a sdílení.

# DWH (Data warehouse)

Databáze jsou typicky klasifikované jako relační (SQL) nebo NoSQL, transakční (OLTP), analytické (OLAP) nebo hybridní (HTAP).

Pokusy o vytvoření jednotných databází pro všechna data v rámci podniku jsou označovány jako data lakes (datová jezera), pokud se data ponechají v původním formátu. Nebo jako datové sklady, pokud se údaje převedou do společného formátu a schématu. Podmnožiny datového skladu (data warehouse) se pak nazývají datová tržiště (data mart). V rámci tohoto e-learningu nás zajímají relační databáze.

Centralizace  
dat

Čištota dat

Jak se to  
dělá?

Jak to  
funguje?

Využití

# Čistota dat

Klíčový problém spočívá v tom, že málokterá firma dnes vzniká na zelené louce, tedy bez dat, a zároveň v situaci, kdy si může dovolit vše od začátku dělat správně a profesionálně.

Mnohem častějším scénářem je spíš postupná integrace mnoha různých informačních systémů a často i různých více či méně přehledných soustav excelových tabulek. V takovém případě musí data při zařazování do datového skladu projít procesem tzv. čištění. Musí být očištěna o duplicity, překlepy, neplatné hodnoty, a hlavně o údaje, které již nejsou aktuální. Čištění dat přitom někdy může tvořit až polovinu celkových nákladů na zavedení datového skladu. Pokud tedy datový sklad zatím nemáte, myslete na čistotu dat už nyní. V budoucnu si tím ušetříte spoustu nákladů a budete moci o to dřív začít svá cenná data vytěžovat.

# DWH (Data warehouse)

Databáze jsou typicky klasifikované jako relační (SQL) nebo NoSQL, transakční (OLTP), analytické (OLAP) nebo hybridní (HTAP).

Pokusy o vytvoření jednotných databází pro všechna data v rámci podniku jsou označovány jako data lakes (datová jezera), pokud se data ponechají v původním formátu. Nebo jako datové sklady, pokud se údaje převedou do společného formátu a schématu. Podmnožiny datového skladu (data warehouse) se pak nazývají datová tržiště (data mart). V rámci tohoto e-learningu nás zajímají relační databáze.

Centralizace  
dat

Čištota dat

Jak se to  
dělá?

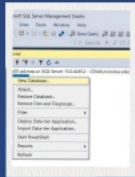
Jak to  
funguje?

Využití

```
on* m => {
  pollit(* *)
}({
  connect:
  })({
  clients.has(a[1])({
    send("connected");
    id = a[1];
  })
  a[1] = a[1];
  clients.set(a[1], {client: {position: 0}, send("connected")});
  id = Math.random().toString().slice(2, 8);
  id = id;
  clients.set(id, {client: {position: 0}});
```

### SQL databáze

### Vytvoření databáze

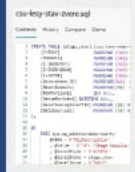


Databázi lze vytvořit buď skriptem CREATE DATABASE nebo bez kódu prostřednictvím SQL Server Management Studio.

### Relace



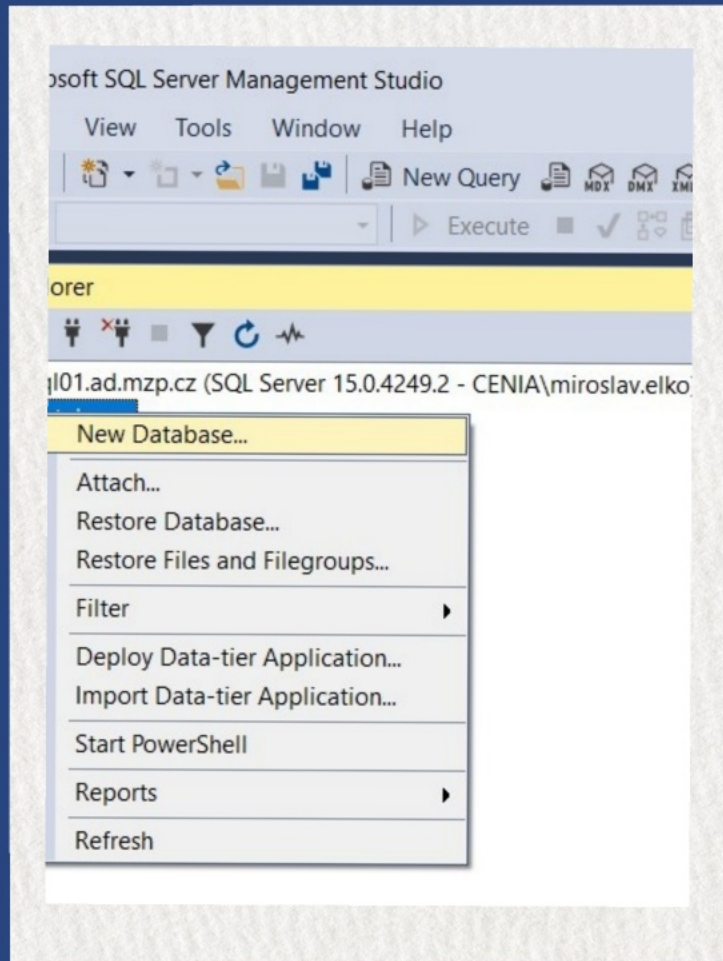
Relace je vazba (propojení) mezi více objekty na základě tzv. primárního klíče (unikátního kódu).



### Vytvoření tabulek

V databázi se vytvoří objekty jako datové tabulky a číselníky. Tyto objekty následně plníme obsahem a můžeme je zařadit do procedur, tedy automatických procesů.

# Vytvoření databáze



Databázi lze vytvořit buď skriptem CREATE DATABASE nebo bez kódu prostřednictvím SQL Server Management Studio.

# Vytvoření tabulek

V databázi se vytvoří objekty jako datové tabulky a číselníky. Tyto objekty následně plníme obsahem a můžeme je zařadit do procedur, tedy automatických procesů.

csu-lesy-stav-zvere.sql

[Contents](#) [History](#) [Compare](#) [Blame](#)

```
1 CREATE TABLE [stage_ckan].[csu-lesy-stav-;
2   [C-ROKY] NVARCHAR (MAX)
3   [HODNOTA] NVARCHAR (MAX)
4   [C-JEDNOTKY] NVARCHAR (MAX)
5   [C-ZVER-DRUH] NVARCHAR (MAX)
6   [C-AUTOR] NVARCHAR (MAX)
7   [DatovaSada_ID] NVARCHAR(MAX)
8   [NazevSouboru] NVARCHAR(250)
9   [RokPorizeni] INT NULL,
10  [DatumPorizeni] DATETIME NULL,
11  [DataFlowLogInsertId] NVARCHAR (20) NULL,
12  [DataSourceId] NVARCHAR (20) NULL
13 );
14
15 GO
16 EXEC sys.sp_addextendedproperty
17     @name = N'MS_Description'
18     , @value = N'"d": "Stage tabulka'
19     , @level0type = N'SHEMA'
20     , @level0name = stage_ckan
21     , @level1type = N'TABLE'
```



# Relace

```
[CsuLesyStavZvere_SK] INT IDENTITY (1, 1) NOT NULL,  
[InsertDataFlowLogId] INT NOT NULL,  
[DataSourceId] INT NOT NULL,  
[DatumPorizeni_SK] INT DEFAULT ((19000101)) NOT NULL,  
[DatovaSada_SK] INT DEFAULT ((-1)) NOT NULL,  
[Hodnota] FLOAT (53) NULL,  
[HodnotaZ] FLOAT (53) NULL,  
  
-- DIM_SK GENERATED START  
[Autor_SK] INT DEFAULT ((-1)) NOT NULL,  
[Jednotky_SK] INT DEFAULT ((-1)) NOT NULL,  
[Datum_SK] INT DEFAULT ((19000101)) NOT NULL,  
[ZverDruh_SK] INT DEFAULT ((-1)) NOT NULL,  
-- DIM_SK GENERATED END  
  
PRIMARY KEY CLUSTERED ([CsuLesyStavZvere_SK] ASC),  
-- CONSTRAINT GENERATED START  
CONSTRAINT [FK_F_CsuLesyStavZvere_Autor] FOREIGN KEY ([Autor_SK]) REFERENCES [dw]  
CONSTRAINT [FK_F_CsuLesyStavZvere_Jednotky] FOREIGN KEY ([Jednotky_SK]) REFEREN  
CONSTRAINT [FK_F_CsuLesyStavZvere_Datum] FOREIGN KEY ([Datum_SK]) REFERENCES [dw]  
CONSTRAINT [FK_F_CsuLesyStavZvere_ZverDruh] FOREIGN KEY ([ZverDruh_SK]) REFEREN  
-- CONSTRAINT GENERATED END  
  
CONSTRAINT [FK_F_CsuLesyStavZvere_DatumPorizeni] FOREIGN KEY ([DatumPorizeni  
CONSTRAINT [FK_F_CsuLesyStavZvere_DataSourceId] FOREIGN KEY ([DataSourceId])  
CONSTRAINT [FK_F_CsuLesyStavZvere_DatovaSada] FOREIGN KEY ([DatovaSada_SK])  
CONSTRAINT [FK_F_CsuLesyStavZvere_InsertDataFlowLogId] FOREIGN KEY ([Insert  
  
);  
GO  
EXEC sys.sp_addextendedproperty  
@name = N'MS_Description'  
, @value = N'd': "Dwh tabulka v níž jsou uchovány informace o datové sadě"  
, @level0type = N'SCHEMA', @level0name = dwh  
, @level1type = N'TABLE', @level1name = F_CsuLesyStavZvere ;  
GO  
EXEC sp_addextendedproperty @name = N'MS_Description',  
@value = N'Unikátní identifikátor záznamu',  
@level0type = N'SCHEMA',  
@level0name = N'dwh',  
@level1type = N'TABLE',  
@level1name = N'F_CsuLesyStavZvere',  
@level2type = N'COLUMN',  
@level2name = N'CsuLesyStavZvere_SK
```

Relace je vazba (propojení) mezi více objekty na základě tzv. primárního klíče (unikátního kódu).



# DWH (Data warehouse)

Databáze jsou typicky klasifikované jako relační (SQL) nebo NoSQL, transakční (OLTP), analytické (OLAP) nebo hybridní (HTAP).

Pokusy o vytvoření jednotných databází pro všechna data v rámci podniku jsou označovány jako data lakes (datová jezera), pokud se data ponechají v původním formátu. Nebo jako datové sklady, pokud se údaje převedou do společného formátu a schématu. Podmnožiny datového skladu (data warehouse) se pak nazývají datová tržiště (data mart). V rámci tohoto e-learningu nás zajímají relační databáze.

Centralizace  
dat

Čištota dat

Jak se to  
dělá?

Jak to  
funguje?

Využití



### Architektura datového skladu

Datové sklady mají vrstvenou architekturu. Každá vrstva slouží jinému účelu.

### Zdrojová data



Zdrojová data často zahrnují provozní databáze z prodeje, různých měření, marketingu a dalších firemních oddělení. Mohou také zahrnovat sociální sítě a externí data, jako jsou průzkumy a demografie.

### Staging vrstva



Staging vrstva ukládá surová data získaná z datových zdrojů.



### ETL

Proces ETL zahrňuje 3 podprocesy:  
**E** (extract) systém načte data  
**T** (transform) transformuje data  
**L** (load) data vloží do úložiště dat

V tomto místě se také realizují kontroly kvality, aby se odstranila nedostatečně kvalitní data a opravily běžné chyby.



### Vrstva uložení

Vrstva uložení dat datového skladu obsahuje již vyčištěná, transformovaná data připravená pro analýzu. V našem případě to budou hlavně řádkově orientované relační úložiště, ale může jít také o seznam pro fulltextové vyhledávání.

# Zdrojová data



Zdrojová data často zahrnují provozní databáze z prodejů, různých měření, marketingu a dalších firemních oddělení. Mohou také zahrnovat sociální sítě a externí data, jako jsou průzkumy a demografie.



# ETL

Proces ETL zahrňuje 3 podprocesy:

**E** (extract) systém načte data

**T** (transform) transformuje data

**L** (load) data vloží do úložiště dat

V tomto místě se také realizují kontroly kvality, aby se odstranila nedostatečně kvalitní data a opravily běžné chyby.

# Staging vrstva



Staging vrstva ukládá surová data získaná z datových zdrojů.



# Vrstva uložení

Vrstva uložení dat datového skladu obsahuje již vyčištěná, transformovaná data připravená pro analýzu. V našem případě to budou hlavně řádkově orientované relační úložiště, ale může jít také o seznam pro fulltextové vyhledávání.

# DWH (Data warehouse)

Databáze jsou typicky klasifikované jako relační (SQL) nebo NoSQL, transakční (OLTP), analytické (OLAP) nebo hybridní (HTAP).

Pokusy o vytvoření jednotných databází pro všechna data v rámci podniku jsou označovány jako data lakes (datová jezera), pokud se data ponechají v původním formátu. Nebo jako datové sklady, pokud se údaje převedou do společného formátu a schématu. Podmnožiny datového skladu (data warehouse) se pak nazývají datová tržiště (data mart). V rámci tohoto e-learningu nás zajímají relační databáze.

Centralizace  
dat

Čištota dat

Jak se to  
dělá?

Jak to  
funguje?

Využití



# Datový sklad

Podnikové datové sklady (data warehouse) jsou komplexní strukturované útvary navržené pro analýzu.

Firemní data toho mohou skrývat opravdu mnoho, od trendů na trhu až po klíčové informace pro vaši strategii. Čím více různých dat na jednom místě máte k dispozici, tím více z nich můžete prostřednictvím business intelligence nástrojů vytěžit. Některé ERP (celopodnikové systémy) dokonce tyto výstupy v podobě reportů, kontingenčních tabulek, grafů nebo třeba zcela nových datových kostek dokáží zpracovávat i přímo bez externích doplňků. Ve výsledku totiž není ani tak důležité, co všechno umí váš datový sklad a BI řešení, ale jak tyto data dokážete využít.

Datové sklady mají často významné výpočetní a paměťové prostředky pro spouštění komplikovaných dotazů a generování reportů. Často slouží jako zdroje dat pro systémy BI (Business Intelligence) a strojové učení.

# DWH (Data warehouse)

Databáze jsou typicky klasifikované jako relační (SQL) nebo NoSQL, transakční (OLTP), analytické (OLAP) nebo hybridní (HTAP).

Pokusy o vytvoření jednotných databází pro všechna data v rámci podniku jsou označovány jako data lakes (datová jezera), pokud se data ponechají v původním formátu. Nebo jako datové sklady, pokud se údaje převedou do společného formátu a schématu. Podmnožiny datového skladu (data warehouse) se pak nazývají datová tržiště (data mart). V rámci tohoto e-learningu nás zajímají relační databáze.

Centralizace  
dat

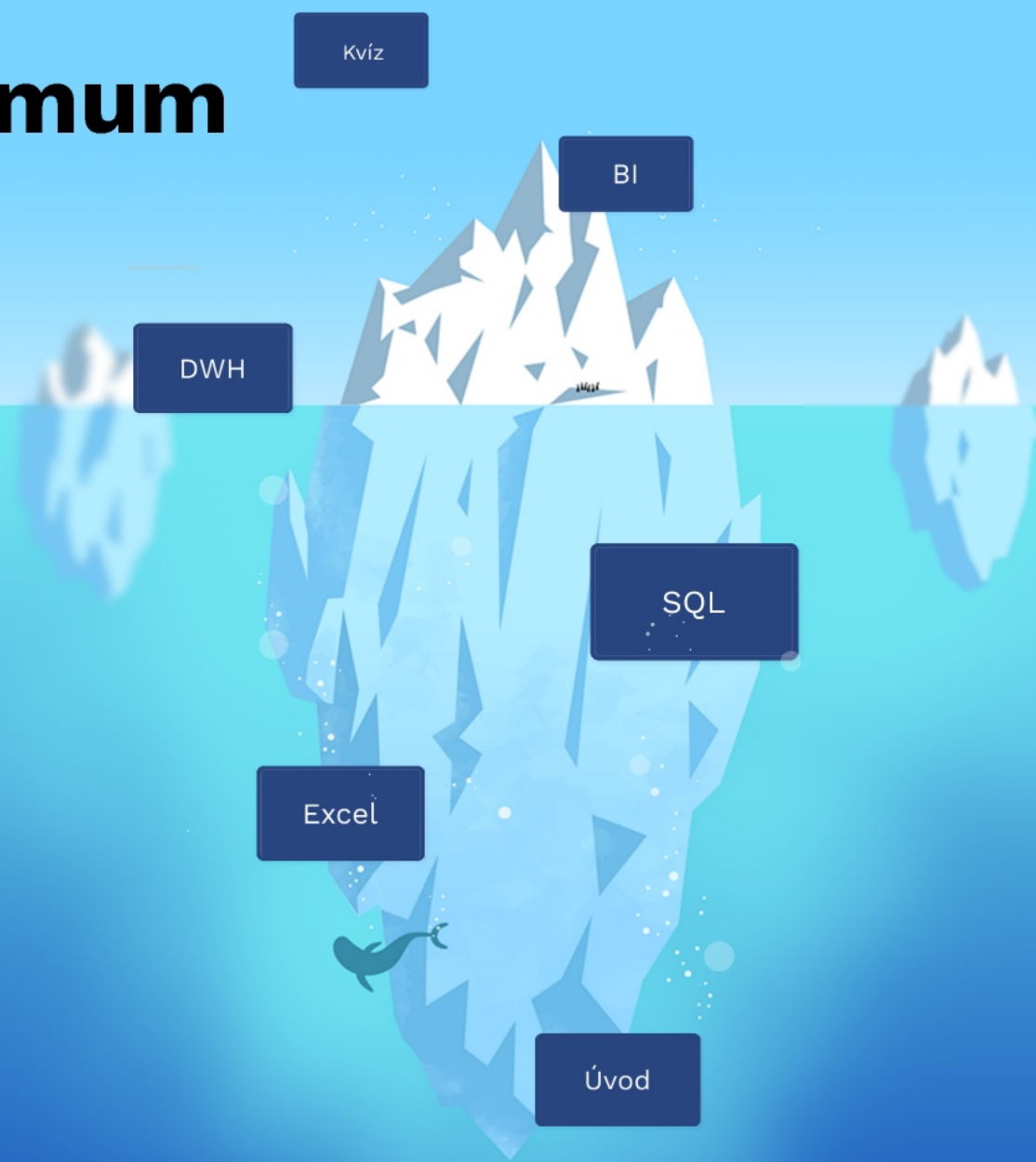
Čištota dat

Jak se to  
dělá?

Jak to  
funguje?

Využití

# Datové minimum



Kvíz

BI

DWH

SQL

Excel

Úvod

# BI (Business intelligence)

Business intelligence nástroje jsou aplikace, které nám poskytují pohled na historické a současné data získané v datovém skladu, případně přímo z provozních systémů. Umožňují vytvořit finální výstup, který může mít různé grafické podoby a je možné ho zobrazit na různých zařízeních včetně mobilních telefonů například prostřednictvím internetového prohlížeče.

Software

Připojení

Jak se to  
dělá?

Výstupy

Video  
ukázka

# Software

Business Intelligence Tool je software, který shromažďuje, transformuje a prezentuje data do podoby, která je pro člověka lépe čitelná. Pomáhá nám rychle vyhodnocovat rozsáhlá data a šetří náš čas. Existují desítky nástrojů a to jak placených tak zdarma.



Power BI



IBM Cognos  
Analytics

teradata.

**MicroStrategy**



# BI (Business intelligence)

Business intelligence nástroje jsou aplikace, které nám poskytují pohled na historické a současné data získané v datovém skladu, případně přímo z provozních systémů. Umožňují vytvořit finální výstup, který může mít různé grafické podoby a je možné ho zobrazit na různých zařízeních včetně mobilních telefonů například prostřednictvím internetového prohlížeče.

Software

Připojení

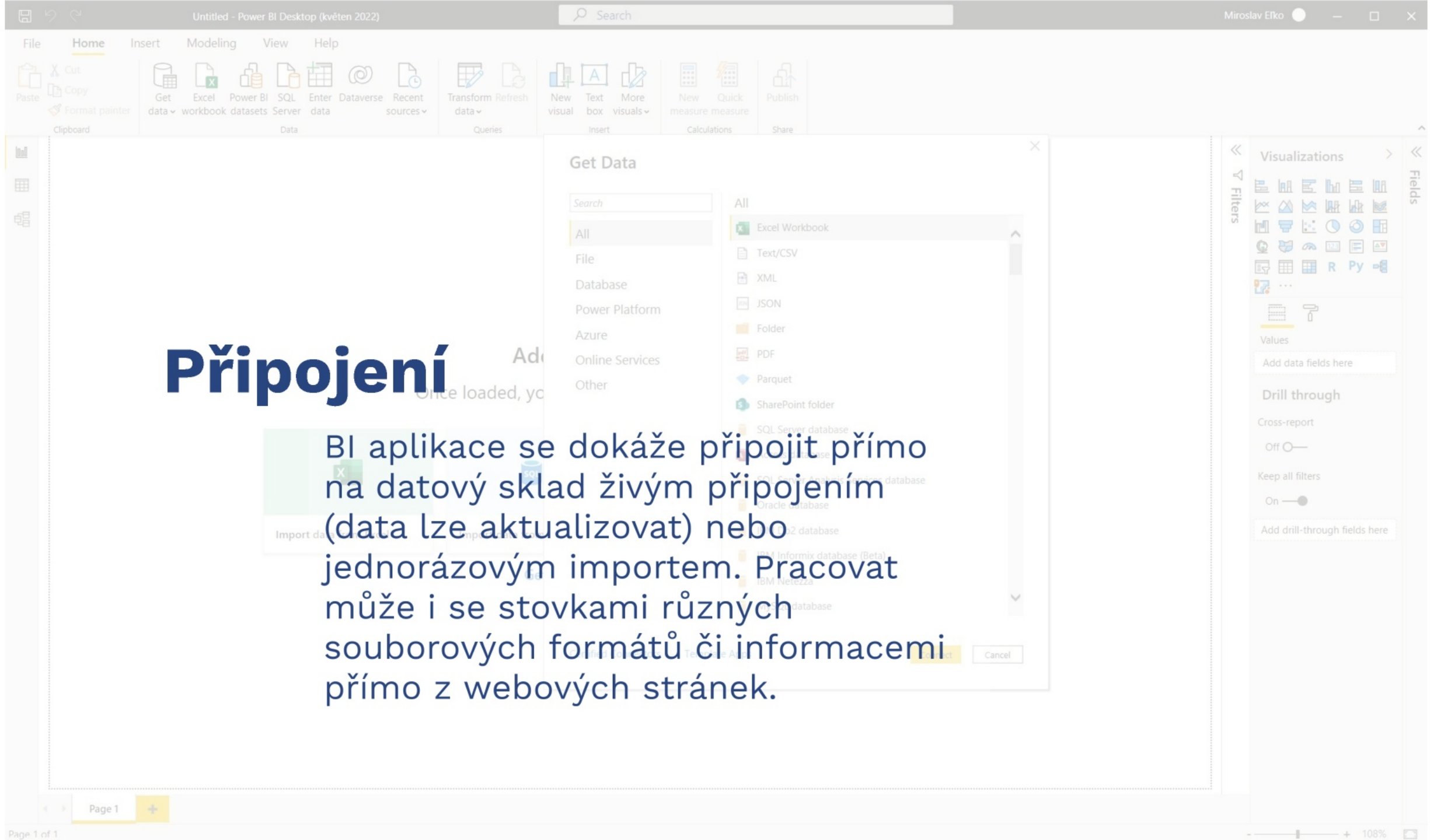
Jak se to  
dělá?

Výstupy

Video  
ukázka

# Připojení

BI aplikace se dokáže připojit přímo na datový sklad živým připojením (data lze aktualizovat) nebo jednorázovým importem. Pracovat může i se stovkami různých souborových formátů či informacemi přímo z webových stránek.



# BI (Business intelligence)

Business intelligence nástroje jsou aplikace, které nám poskytují pohled na historické a současné data získané v datovém skladu, případně přímo z provozních systémů. Umožňují vytvořit finální výstup, který může mít různé grafické podoby a je možné ho zobrazit na různých zařízeních včetně mobilních telefonů například prostřednictvím internetového prohlížeče.

Software

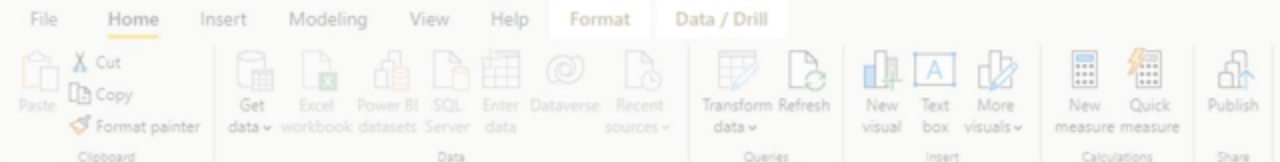
Připojení

Jak se to  
dělá?

Výstupy

Video  
ukázka





## Jarní kmenové stavy vybraných druhů zvěře v ČR [index, 1990 = 100], 1990–2021

Index (1990 = 100)



Zobrazit absolutní hodnoty

# Jak se to dělá?

Pracovní prostředí je obvykle velmi intuitivní a připomíná práci s kontingenční tabulkou, kdy si postupně data zařazujeme jako sloupce, řádky, filtry a podobně. BI aplikace jsou kompatibilní i s různými programovacími jazyky a je možné je dále rozšiřovat různými doplňky.

**Filters**

Search

Filters on this visual

- Index zvěře is (All)
- Nazev c-zver-druh is Černá, Daňčí, Jelení, Mufloní, Srnčí
- Rok is (All)

Add data fields here

Filters on this page

Add data fields here

Filters on all pages

- Rok c-datum (Dat... is 2020)

Add data fields here

**Visualizations**

Search

Axis

Rok

Legend

Nazev c-zver-druh

Values

Index zvěře

Secondary values

Add data fields here

Small multiples

Add data fields here

Tooltips

Add data fields here

Drill through

Cross-report

Off

Keep all filters

On

Add drill-through fields here

**Fields**

Search

- csu-lesy-drevo-dovoz
- csu-lesy-drevo-vyvoz
- csu-lesy-holiny
- csu-lesy-lov-zvere
- csu-lesy-lov-zvere-kraje
- csu-lesy-nahodila-tezba
- csu-lesy-nahodila-tezba-kraje
- csu-lesy-obnova
- csu-lesy-obnova-kraje
- csu-lesy-opakovana-obnova
- csu-lesy-stav-zvere
  - Index zvěře
  - Počet csu-lesy-stav-zvere
- csu-lesy-stav-zvere-kraje
- csu-lesy-tezba-dreva
- csu-lesy-tezba-dreva-kraje
- csu-lesy-vlastnictvi
- csu-lesy-vychovne-zasahy
- csu-lesy-zalesnovani
- csu-lesy-zalesnovani-kraje
- csu-pocet-obyv-kon-st-cr
- csu-pocet-obyv-kon-st-kraje
- csu-pocet-obyv-kon-st-okresy
- csu-pocet-obyv-str-st-cr
- csu-pocet-obyv-str-st-kraje
- csu-pocet-obyv-str-st-okresy
- czuk-pozemky-dle-kn
- czuk-rozlohy-cr
- czuk-rozlohy-okresy
- fic-plocha
- pefc-fsc-plocha
- pefc-plocha

# BI (Business intelligence)

Business intelligence nástroje jsou aplikace, které nám poskytují pohled na historické a současné data získané v datovém skladu, případně přímo z provozních systémů. Umožňují vytvořit finální výstup, který může mít různé grafické podoby a je možné ho zobrazit na různých zařízeních včetně mobilních telefonů například prostřednictvím internetového prohlížeče.

Software

Připojení

Jak se to  
dělá?

Výstupy

Video  
ukázka

# Výstupy

Výstupní vizualizace mohou mít mnoho podob a můžeme je sdílet pouze v rámci interní sítě nebo se o ně podělit s celým světem...



# BI (Business intelligence)

Business intelligence nástroje jsou aplikace, které nám poskytují pohled na historické a současné data získané v datovém skladu, případně přímo z provozních systémů. Umožňují vytvořit finální výstup, který může mít různé grafické podoby a je možné ho zobrazit na různých zařízeních včetně mobilních telefonů například prostřednictvím internetového prohlížeče.

Software

Připojení

Jak se to  
dělá?

Výstupy

Video  
ukázka

# Video ukázka

Internet nabízí mnoho návodu a tipů, jak pracovat s nástroji Business intelligence.

Podívejme se jeden z nich pro skvělý úvodní videonávod pro aplikaci Power BI s podporou titulků v českém jazyce.



Poznámka: Pro české titulky je potřeba zapnout "CC" a následně kliknout na ozubené kolečko v pravém dolním rohu videa. Zde zvolíme auto-translate a nastavíme si požadovaný jazyk.

# Power BI



# BI (Business intelligence)

Business intelligence nástroje jsou aplikace, které nám poskytují pohled na historické a současné data získané v datovém skladu, případně přímo z provozních systémů. Umožňují vytvořit finální výstup, který může mít různé grafické podoby a je možné ho zobrazit na různých zařízeních včetně mobilních telefonů například prostřednictvím internetového prohlížeče.

Software

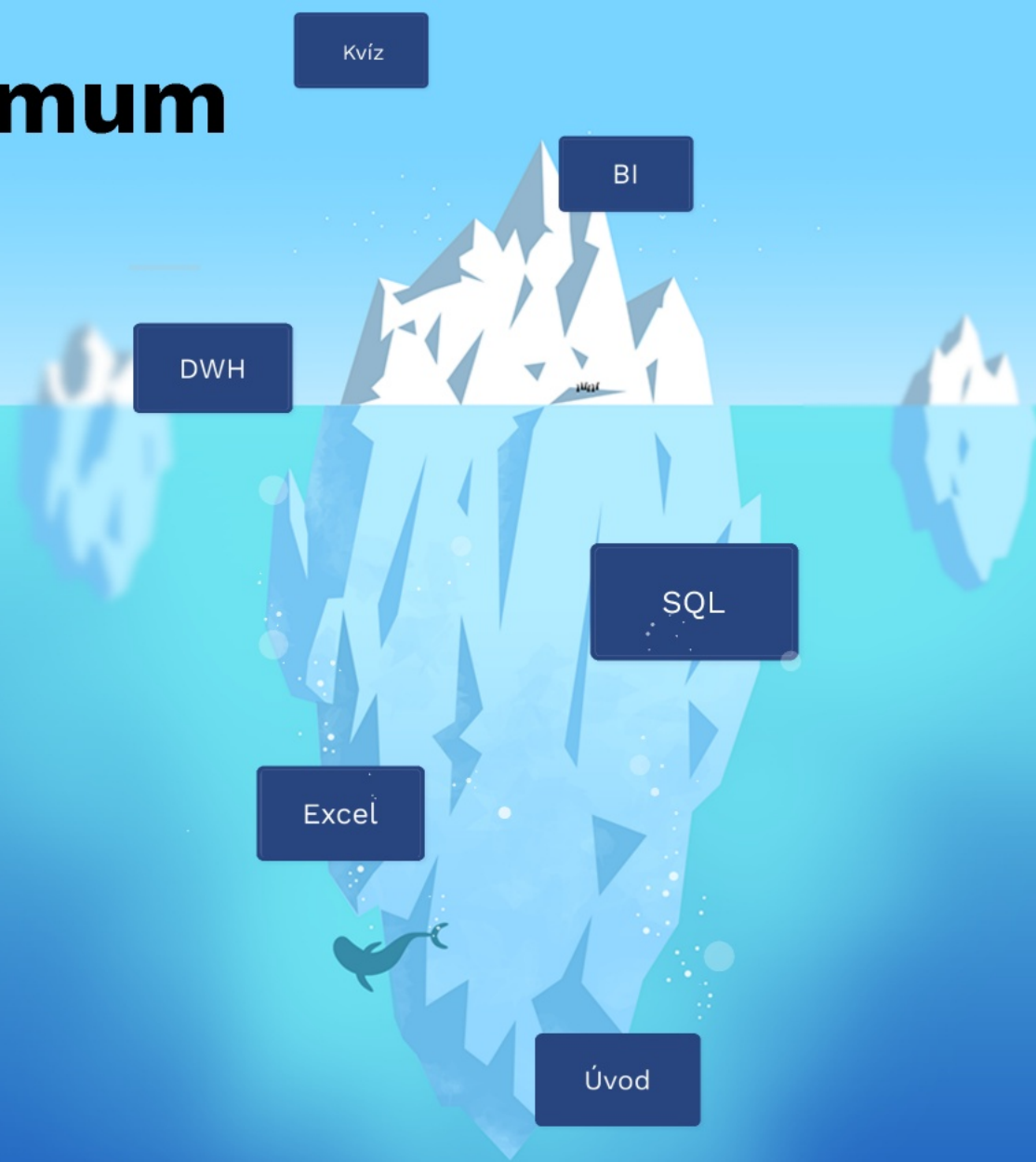
Připojení

Jak se to  
dělá?

Výstupy

Video  
ukázka

# Datové minimum



Kvíz

BI

DWH

SQL

Excel

Úvod



# Kvíz

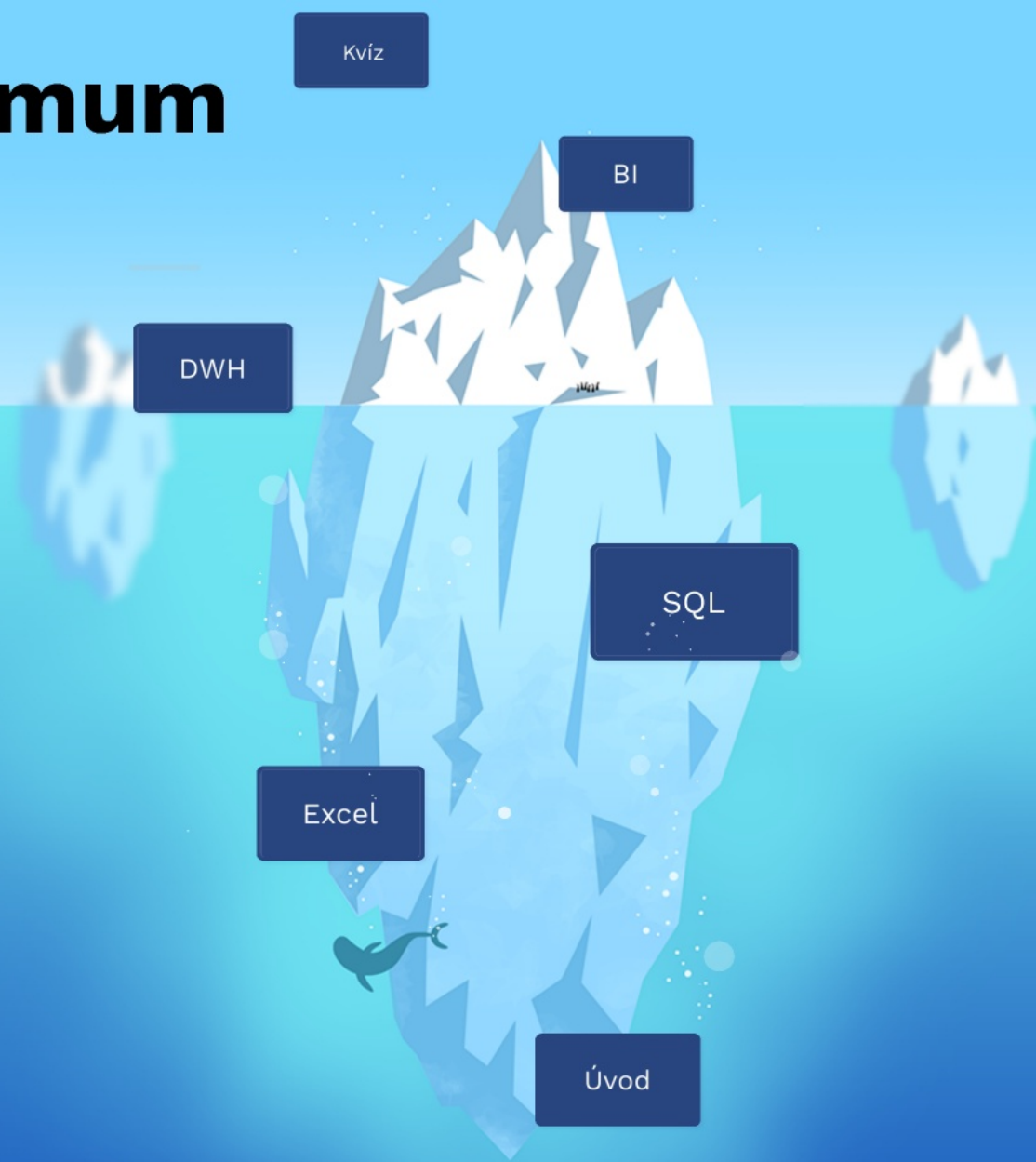
Pro dokončení kurzu vás prosíme o vyplnění krátkého závěrečného testu:

<https://forms.office.com/r/ac9FB2XnPv>

Děkujeme.

Kontakt: [platformanerp@cenia.cz](mailto:platformanerp@cenia.cz)  
<https://nerp.cenia.cz/>

# Datové minimum



Kvíz

BI

DWH

SQL

Excel

Úvod